UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

In Re: OpenAI, Inc. Copyright Infringement
Litigation

This Document Relates To:

    Case No. 1:23-cv-08292
    Case No. 1:23-cv-10211
    Case No. 1:23-cv-00084
    Case No. 1:25-cv-03291
    Case No. 1:25-cv-03297
    Case No. 1:25-cv-03482
    Case No. 1:25-cv-03483

Case No. 1:25-md-3143 (SHS) (OTW)

**DEFENDANT MICROSOFT CORPORATION'S REPLY BRIEF IN SUPPORT OF
ITS MOTION TO STRIKE PORTIONS OF THE CONSOLIDATED CLASS
ACTION COMPLAINT**

TABLE OF CONTENTS

<u>**TABLE OF AUTHORITIES**</u>

Page(s)

Cases

INTRODUCTION

Plaintiffs do not seriously dispute that the material that Microsoft seeks to strike from the Consolidated Class Action Complaint ("CCAC") exceeds the scope of the previous allegations asserted against Microsoft and would require significant additional discovery that could not be accomplished under the current case schedule. But instead of providing a clean and unambiguous complaint that would resolve Microsoft's concerns and ease the administration of this case going forward, Plaintiffs hide behind three unpersuasive arguments: (1) that Microsoft's concerns can be resolved by stipulation; (2) that the words used in the CCAC do not mean what they literally say, and that Microsoft should instead take Plaintiffs' word regarding they really meant; and (3) that by consolidating these actions, Microsoft should face the most expansive set of allegations from the underlying cases even though it was previously not a party to several of those cases. None of these arguments have merit.

As set forth in Microsoft's opening brief, a reasonable reading of the CCAC demonstrates that this pleading on its face would dramatically expand the scope of the case by adding new allegations regarding: (1) Microsoft's own trained models; (2) OpenAI models developed and released after Plaintiffs filed suit; and (3) copyrighted works beyond just books. Each of these sets of allegations fall outside the scope of discovery conducted to date, and adding them at this late stage would serve little purpose but to delay the case schedule—an outcome the Court has made clear it does not want, and which Plaintiffs themselves should not be in favor of. Yet Plaintiffs seek to skirt around existing Orders and guidance from the Court warning against this very type of case expansion. The impermissible portions of the CCAC should accordingly be stricken and the CCAC amended to ensure that it does not encompass allegations that go beyond those "products and causes of action" that were initially pled by Plaintiffs.

1

ARGUMENT

Microsoft was forced into bringing this Motion by the Plaintiffs' decision to introduce into the CCAC a number of ambiguous allegations in contravention of the Court's guidance against the addition of any expanded allegations. This case was centralized for the primary purpose of efficiency, and it would be contrary to this goal, and fundamentally unfair to Microsoft, to introduce brand new allegations against Microsoft that require significant additional discovery—particularly at this late stage. While Microsoft did not oppose centralization of these cases, it was not the party that sought it. As such, the JPML's recognition of the overlapping issues and decision to centralize these cases does not preclude Microsoft from challenging the propriety of brand-new allegations asserted against it. What is more, the central question at issue in these cases focuses on the *training* of *OpenAI's* large language models and whether or not that constitutes fair use or copyright infringement. The legality of that question can be determined by focusing on the models themselves, with no need to take a frolic and detour into the Microsoft products that utilize those models. Microsoft's Motion should be granted and the claims against it should be limited to only those that were pled originally and have been the subject of discovery for the past 18 months.

I. **ALLEGATIONS IMPLICATING NON-OPENAI-TRAINED MODELS SHOULD BE STRICKEN BECAUSE PLAINTIFFS ADMIT THAT ONLY OPEN-AI-TRAINED MODELS HAVE BEEN AT ISSUE**

In response to Microsoft's Motion, Plaintiffs agree that Microsoft's own trained models are not in the case. ECF 453[1], Memorandum of Law in Opposition to Microsoft's Motion to Strike ("Opposition" or "Opp.") at 1, 4-5. Plaintiffs also purport to represent that they did not

---

[1] All ECF references herein are to the MDL docket, *In re OpenAI, Inc., Copyright Infringement Litigation*, 1:25-md-03143-SHS-OTW.

2

intend to include them in the CCAC. *Id.* at 4-6. Yet, as explained in Microsoft's Motion, a

reasonable reading of the CCAC clearly includes Microsoft's own trained models:

> 294. The Class consists of:
>
> All legal or beneficial owners of copyrighted works that: (A) were registered with the United States Copyright Office within five years of the work's first publication; (B) *were downloaded or otherwise reproduced by OpenAI or Microsoft*; (C) were registered with the United States Copyright Office before being downloaded or otherwise reproduced by OpenAI or Microsoft, or were registered within three months of first publication; and (D) are assigned one or more International Standard Books Number(s) (ISBN) or Amazon Standard Identification Number(s) (ASIN).
>
> 311. Defendants infringed on the rights, under 17 U.S.C. § 106, of Plaintiffs and members of the proposed Class by, among other things, *reproducing the works owned by Plaintiffs and the proposed Class from shadow libraries and to create and develop datasets used to train their artificial intelligence models*. Specifically, Plaintiffs and members of the Class assert claims for direct infringement against *both Microsoft and OpenAI*.

CCAC ¶¶ 294, 311 (emphasis added). Moreover, these allegations materially differ from the

allegations in the underlying Complaints, which were more clearly focused only on issues related

to OpenAI's models. Given the Court's prior rulings, and Plaintiffs' own representations, these

expanded allegations cannot stand.

Yet, even though they agree that these models are not in the case, Plaintiffs are refusing to

amend the CCAC to make this clear. Opp. at 4-5. It is the pleadings that control, however, not

Plaintiffs' representations about what they intended the pleadings to mean, and the pleadings

should correctly reflect the claims asserted. *See, e.g., Siti-Sites.com, Inc. v. Verizon Commc'ns,

Inc.*, No. 10-cv-3751, 2010 WL 5392927, at *5 (S.D.N.Y. Dec. 29, 2010), *aff'd*, 428 F. App'x

100 (2d Cir. 2011) ("A party may not amend its complaint, however, through statements made in

its motion papers."). Plaintiffs' representation that they will stipulate that this case does not

include "Microsoft-trained LLM[s]" is similarly insufficient.[2] *See id.* Plaintiffs should either

voluntarily amend the pleading to accurately reflect the claim they now say they meant to bring,

or these portions of the CCAC should be stricken and replaced with a version that accurately

reflects Plaintiffs' claims.

**II.   ALLEGATIONS REGARDING GPT MODELS RELEASED BEFORE GPT-3
AND AFTER GPT-4 TURBO, AND THE PRODUCTS THOSE MODELS
UNDERLIE, SHOULD BE STRICKEN BECAUSE THEY DO NOT IMPACT
MICROSOFT'S ALLEGED LIABILITY AND WOULD ONLY DELAY
DISCOVERY**

As explained in its Motion, Microsoft further seeks to strike references to earlier and later

released GPT models that were not previously asserted against Microsoft because including these

models runs directly afoul of the numerous orders issued by this Court limiting the scope of

models and products in this case. *See* ECF 327, Defendant Microsoft Corporation's

Memorandum of Law in Support of Its Motion to Strike Portions of the Consolidated Class

Action Complaint ("Motion" or "Mot.") at 14.

For starters, Plaintiffs agree that the older models, GPT 1 and 2, are not asserted against

Microsoft. Opp. at 7. So, with respect to those models there is no debate that the CCAC should

be clarified.

Plaintiffs' primary argument for including the newer models, and the products they

underlie, is that by centralizing these cases, Plaintiffs are entitled to allege the ***most expansive***

allegations contained in all of the underlying cases against all the parties. Opp. at 7-8. Since

these models were asserted in the *Tremblay* case in the Northern District of California, Plaintiffs

argue that they may be asserted here. *Id.* But Microsoft was never a defendant in that case, and

---

[2] Even if Plaintiffs' representation that they "will stipulate" to clarifying their CCAC was sufficient (Opp. at 4)—it is not—Plaintiffs do not provide in connection with their Opposition (nor have they provided outside the briefing before this Court) a stipulation for Defendants' consideration, rendering this representation hollow.

thus these newer models and the related products they power should remain outside the scope of

the case *against Microsoft* and should be stricken from the CCAC, or their lack of applicability

to Microsoft otherwise clarified. Microsoft continues to be an add-on in these cases brought

against OpenAI's technology and LLM training, but it should not be deprived of due process by

unfairly facing significantly expanded allegations against it within the confines of the Court's

established case schedule as a result of Plaintiffs' deliberate choices in how they are litigating

these cases. Centralization of these cases does not give Plaintiffs the right to circumvent the

Court's orders and to assert brand new models and products *for the first time against Microsoft*

at this late stage.[3] *See In re Digital Advert. Antitrust Litig.*, 555 F. Supp. 3d 1372, 1376 (J.P.M.L.

2021) ("[T]he transferee court may account, at his discretion, for any differences among the

actions by using appropriate pretrial devices, such as separate tracks for discovery or motion

practice.").

Importantly, keeping these models and products out of this case will not meaningfully

impact the substance of Plaintiffs' claims. Because Plaintiffs allege that the infringement by

these later-released models is the same as what is already asserted against Microsoft, their

inclusion will not make any difference to Plaintiffs' liability arguments. In other words, the

question of whether training GPT 3 on Plaintiffs' copyrighted works is infringement or fair use is

no different than the question of whether training GPT 4o on Plaintiffs' copyrighted works is

infringement or fair use. Assuming *arguendo* that there is some baseline evidence of copying

and use with respect to the training of each model, the legal analysis is not going to materially

---

[3] As noted in Microsoft's Motion, although OpenAI agreed to provide discovery into certain
models, including GPT-1, GPT-2, GPT-4o, and GPT-4o Mini, Microsoft **never agreed** that these
models were asserted against it, nor were these models alleged in the underlying complaint against
it, and discovery for the past 18 months has not included them as to Microsoft. Mot. at 14, n.8.

change on a fundamental level from model to model, such that if Defendants' fair use defense is successful, Plaintiffs' alleged infringement claims based on any of the models would be meritless.[4]

On the flip side, allowing Plaintiffs to maintain an amended Complaint that includes Microsoft products that are powered by these models would dramatically increase the scope of discovery and do nothing more than delay the progression of this case—an outcome the Court has made clear is not desirable. Mot. at 14-16; *see also* ECF 238, Joint Case Schedule ("Scheduling Order") (noting that "[t]here will be no adjournments of" the deadline to file Summary Judgment). The products that use these later-released models were not in existence at the time the first lawsuits underlying this MDL were filed. Inclusion of such products would only serve to exponentially increase discovery in this case, rendering it impossible to complete within the current schedule. This is a frolic into products that were not previously accused, would require discovery exponentially larger that the discovery already contemplated and completed, and would have no real impact on the evaluation of the issues of liability before the Court. Therefore, their inclusion is not proportional to the needs of the case, in particular as to Microsoft, and the allegations asserting infringement based on models other than GPT-3, GPT-3.5, GPT-4, and GPT-4 Turbo against Microsoft should be stricken.[5]

To the extent the Court is inclined to allow the inclusion of newer models in the case because they were at issue in *Tremblay*, which it should not, it is certainly possible to do so *without* asserting those models against Microsoft or allowing the addition of allegations relating to *Microsoft's products* that run on these newer models. To date, those products themselves

---

[4] It is worth noting, as well, that the later models Plaintiffs seek to add are even less pertinent given that OpenAI has indicated that the books1 and books2 collections were not used to train any model past GPT 3.5.

have not been separately accused of infringement, rather Plaintiffs' claims have focused on the

training of the models and not their implementation or subsequent use. As such, any Microsoft

products that run on the newer models that were accused in *Tremblay*, but never in the SDNY

matters against Microsoft, are irrelevant and should not be allowed into the case at this time.

III.   THE CLASS DEFINITION SHOULD BE STRICKEN TO THE EXTENT NON-BOOKS COULD FALL WITHIN THE DEFINITION

Finally, as set forth in Microsoft's Motion, Microsoft also seeks to strike portions of the

CCAC that sweep in copyrighted works beyond just fiction and nonfiction books into the class

definition. Mot. at 17-18. Yet again relying on the mantra of "do what I meant, not what I said,"

Plaintiffs point to other allegations to argue that "no one could be confused that this case is about

books." Opp. at 9. However, the class definition *is* confusing. As currently pleaded, Plaintiffs

define the class to include "[a]ll legal or beneficial owners of copyrighted works that… are

assigned one or more International Standard Books Number(s) (ISBN) **or Amazon Standard**

Identification Number(s) (ASIN)." CCAC ¶ 294 (emphasis added). Despite Plaintiffs'

arguments to the contrary, the definition does *not* expressly limit the potential class works to

books, and there are numerous other copyrightable works beyond books that are sold on Amazon

and assigned an ASIN number. Mot. at 18. Plaintiffs' reliance on *other* allegations to suggest

how its class definition should be read and interpreted, rather than the language of the definition

itself, does nothing but illustrate the inconsistency and ambiguity of the CCAC—a problem that

must be rectified now, not later as Plaintiffs argue.

In opposing Microsoft's Motion, Plaintiffs argue that they need not clarify the definition

at this stage because it is an issue that "can come later," relying on the class definition adopted

by Judge Alsup in the *Bartz* case as reason why this inconsistent definition does not matter at this

stage. Opp. at 10. This is plainly untrue. Expanding the class definition beyond books would

expand the scope of discovery in this case. Given the case schedule, all fact discovery must be

completed before February 27, 2026, which inherently encompasses discovery that would relate

to class certification. *See* ECF 238, Scheduling Order. There is no separate time period built

into the case schedule that would support Plaintiffs' argument that clarification of this definition

can wait. It cannot. The differences between the class definition in the CCAC and the

underlying Complaints in the matters prior to consolidation are not just a matter of semantics.

The CCAC's class definition is a major departure from the prior class definitions and, as

currently read, would rope entirely new industries into this case necessitating extensive new

discovery and potentially an entirely different roster of expert witnesses. If Plaintiffs will not

agree to voluntarily amend the CCAC to make sure it aligns with their intentions, then striking

the class definition is warranted.

CONCLUSION

For the reasons presented herein, Microsoft respectfully requests the Court strike

paragraphs 294 and 311 of the CCAC subject to limiting amendments as described herein,

references in paragraphs 4, 5, 81-82, 89-90, 93, 96-97, 117, 127, and 161 to models other than

those released between GPT-3 and GPT-4 Turbo, and reference to ASIN numbers in paragraph

294.

Dated: August 28, 2025   */s/ Laura B. Najemy*

                                                Annette L. Hurst (admitted *pro hac vice*)
                                                ORRICK, HERRINGTON &
                                                SUTCLIFFE LLP
                                                The Orrick Building
                                                405 Howard Street
                                                San Francisco, CA 94105
                                                Telephone: (415)773-5700
                                                ahurst@orrick.com

8

Christopher Cariello
Marc Shapiro
ORRICK, HERRINGTON &
SUTCLIFFE LLP
51 West 52nd Street
New York: NY 10019
Telephone: (212) 506-3778
Facsimile: (212) 506-5151
ccariello@orrick.com
mrshapiro@orrick.com

Sheryl Koval Garko (admitted *pro hac vice*)
Laura Brooks Najemy (admitted *pro hac vice*)
ORRICK, HERRINGTON &
SUTCLIFFE LLP
222 Berkeley Street, Suite 2000
Boston, MA 02116
Telephone: (617) 880-1800
Facsimile: (617) 8801-1801
sgarko@orrick.com
lnajemy@orrick.com

Jeffrey S. Jacobson
FAEGRE DRINKER BIDDLE
& REATH LLP
1177 Avenue of the Americas
New York, New York 10036
Telephone: (212) 248-3191
jeffrey.jacobson@faegredrinker.com

Jared B. Briant (admitted *pro hac vice*)
Kirstin L. Stoll-DeBell (admitted *pro hac vice*)
FAEGRE DRINKER BIDDLE
& REATH LLP
1144 Fifteenth Street, Suite 3400
Denver, CO 80202
Telephone: (303) 607-3500
jared.briant@faegredrinker.com
kirstin.stolldebell@faegredrinker.com

Carrie A. Beyer (admitted *pro hac vice*)
FAEGRE DRINKER BIDDLE & REATH LLP
320 South Canal Street, Suite 3300
Chicago, IL 60606-5707
Telephone: (312) 569-1000
Facsimile: (312) 569-3000

9

carrie.beyer@faegredrinker.com

Elizabeth M.C. Scheibel (admitted *pro hac vice*)
FAEGRE DRINKER BIDDLE & REATH LLP
2200 Wells Fargo Center, 90 S. 7th Street
Minneapolis, MN 55402
Telephone: (612) 766-7000
Facsimile: (612) 766-1600
elizabeth.scheibel@faegredrinker.com

Attorneys for Defendant Microsoft Corporation

PROOF OF SERVICE VIA ECF

On August 28, 2025, I caused the foregoing document to be served on all counsel of

record via ECF.

/s/ Laura B. Najemy
Laura B. Najemy

11

12

CERTIFICATE OF WORD COUNT

Pursuant to Local Civil Rule 7.1, I hereby certify that the foregoing document was typed using 12-point, Times New Roman font and contains 2,492 words, exclusive of the case caption, table of contents, table of authorities, signature blocks, or any required certificates, but does include material contained in footnotes or endnotes.

/s/ Laura B. Najemy
Laura B. Najemy